yang_seonhyeHW3

Seonhye Yang

January 23, 2019

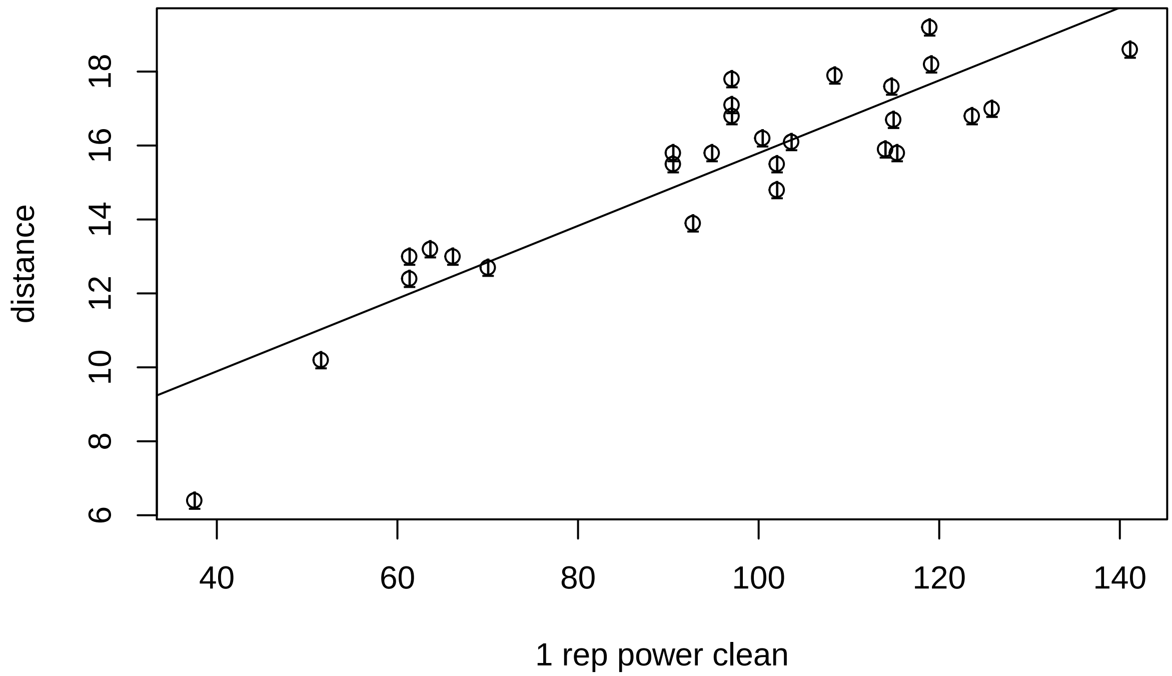
```
library(data.table)
shot_power <- fread("http://www.stat.ufl.edu/~winner/data/shotputt_powerclean.csv")
powershot <- data.table("shotputt_powerclean.csv", header=TRUE)
```

Question 1

```
attach(shot_power, warn.conflicts = FALSE)
model1 <- lm(shot.putt~power.clean, data=powershot)
summary(model1)
```

```
##
## Call:
## lm(formula = shot.putt ~ power.clean, data = powershot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2475 -1.1798  0.3635  0.9516  2.3010
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.959629   0.958835   6.215 1.42e-06 ***
## power.clean 0.098344   0.009721  10.117 1.66e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 26 degrees of freedom
## Multiple R-squared:  0.7974, Adjusted R-squared:  0.7896
## F-statistic: 102.4 on 1 and 26 DF,  p-value: 1.663e-10
```

```
plot(power.clean, shot.putt, xlab="", ylab="")
text(power.clean, shot.putt, rownames(powershot), cex=0.8)
title(xlab="1 rep power clean", ylab="distance")
abline(model1)
```



With an intercept of 5.95, we can say that even with no power cleaning ablity, anyone should be able to have a shot putt distance of 5.95. In addition, with a slope of .098, for every 1 kg increase in power cleaning, there is a .098 foot increase in shot putt distance. This relationship has a strong positive correlation of 0.78, Showing that as your power clean weight goes up, your shot putt distance increases.

Question 2

```
fit1 = lm(shot.putt~I(power.clean-100), data=powershot)
fit1
```

```
##
## Call:
## lm(formula = shot.putt ~ I(power.clean - 100), data = powershot)
##
## Coefficients:
##          (Intercept)  I(power.clean - 100)
##             15.79401               0.09834
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = shot.putt ~ I(power.clean - 100), data = powershot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2475 -1.1798  0.3635  0.9516  2.3010
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          15.794011   0.243083   64.97  < 2e-16 ***
## I(power.clean - 100)  0.098344   0.009721   10.12 1.66e-10 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 26 degrees of freedom
## Multiple R-squared:  0.7974, Adjusted R-squared:  0.7896
## F-statistic: 102.4 on 1 and 26 DF,  p-value: 1.663e-10
```

When centering the model around 100kg, we can see that there is no change in the slope of the model. This means even now, for every 1kg increase in power cleaning weight, there is a .098 foot increase in the shot putt distance. However, our intercept has changed to 15.79, indicating that we are now setting the "base value" of our data to be 100kg of power cleaning weight. This allows us to create a model that would essentially extrapolate further out on the side of people lifting more than 100kg

Question 3

```r
predict(model1, data.frame(power.clean=90))
```

```
##        1
## 14.81057
```

The predicted shot put for an athlete who power cleans 90kg is 14.81

```r
predict(model1, data.frame(power.clean=110))
```

```
##        1
## 16.77745
```

The predicted shot put for an athlete who power cleans 90kg is 14.81

Question 4

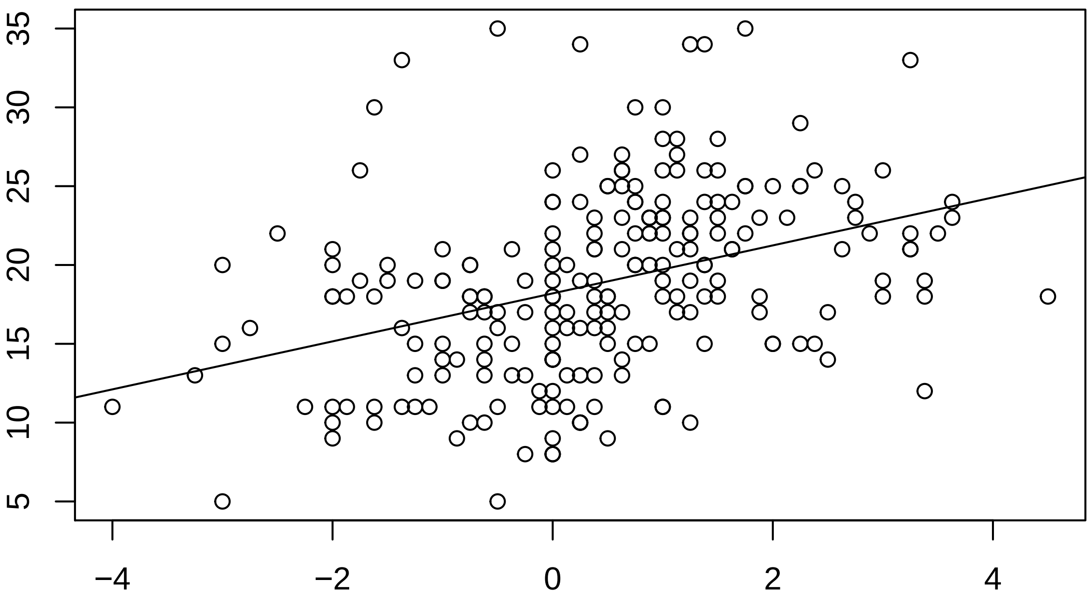
```r
library(data.table)
nfl <- read.csv("nfl_combine2017.csv")
attach(nfl, warn.conflicts = FALSE)
model = lm(BenchPress~I(ArmLength-32), nfl)

summary(model)
```

```
##
## Call:
## lm(formula = BenchPress ~ I(ArmLength - 32), data = nfl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4362  -4.1972  -0.1121   3.3220  17.5638
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        18.1972     0.3790  48.018  < 2e-16 ***
## I(ArmLength - 32)   1.5220     0.2496   6.099 4.56e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.503 on 227 degrees of freedom
```

```
##   (91 observations deleted due to missingness)
## Multiple R-squared:  0.1408, Adjusted R-squared:  0.137
## F-statistic:  37.2 on 1 and 227 DF,  p-value: 4.556e-09
```

```
plot(I(ArmLength-32), BenchPress, xlab="", ylab="")
abline(model)
```



A person with an arm length of 0 can bench press 18 times. Although that is not possible, this data point exists because of the data being centered to have a baseline of 18 reps, due to NFL players being in the data. In addition, with an inch increase, the bench press number increases by 1.52. This has a weak positive correlation of 0.1408 which means as the arm length increases, the number of bench press increases as well.

Question 5

```
attach(nfl, warn.conflicts = F)
predict(model,data.frame(ArmLength=30))
```

```
##        1
## 15.15323
```

We can see from the R^2 value that there is a weak positive correlation. Combining this with the standard error on the predicted value of .25, we have an error that is relatively significant. Finally, our model does visually show linearity, but with the amount of outliers, I would say the regression is not adequate